Revue d'Ecologie (Terre et Vie), Vol. 70 (suppt 12 « *Espèces invasives* »), 2015 : 127-138

MULTIPLE INTRODUCTIONS AND ADMIXTURE AT THE ORIGIN OF THE CONTINENTAL SPREAD OF THE FUNGAL BANANA PATHOGEN *MYCOSPHAERELLA FIJIENSIS* IN CENTRAL AMERICA: A STATISTICAL TEST USING APPROXIMATE BAYESIAN COMPUTATION

Stéphanie ROBERT[1], Marie-Françoise ZAPATER[1], Jean CARLIER[1], Catherine ABADIE[1] & Virginie RAVIGNE[1*]

[1] CIRAD. UMR BGPI. F-34398 Montpellier. France

* Corresponding author: CIRAD - UMR BGPI - Campus International de Baillarguet - TA A 54 / K - 34398 Montpellier Cedex 5 – France. Phone: + 33 4 99 62 48 58 ; Fax: + 33 4 99 62 48 48 ; E-mail: virginie.ravigne@cirad.fr

RÉSUMÉ.— *Introductions multiples et mélange à l'origine de l'expansion continentale du champignon pathogène du bananier* Mycophaerella fijiensis *en Amérique centrale : un test statistique utilisant le calcul bayésien approché.*— La reconstruction des histoires d'invasions est l'un des principaux défis de la biologie des invasions. L'ascomycète *Mycophaerella fijiensis*, agent causal de la maladie des raies noires des bananiers, en provenance d'Asie du Sud-Est, a commencé à envahir le continent américain en 1972 et est aujourd'hui présent dans toutes les régions productrices de bananes d'Amérique. Dans une précédente étude de la structure génétique des populations à l'échelle mondiale, nous supposions que l'invasion américaine pourrait avoir résulté d'introductions multiples et de mélange. Ici, nous avons cherché (i) à tester statistiquement cette hypothèse, et (ii) à inférer les caractéristiques démo-génétiques de cette introduction, en utilisant une technique de choix de scénario par calcul bayésien approché (ABC). Pour ce faire, nous avons confronté sept scénarios d'introduction de la maladie à des données précédemment acquises de génotypage à 21 locus microsatellites de six populations américaines, deux populations d'Asie du Sud-Est et une population océanienne. Nous avons pu démontrer fermement une contribution de l'Océanie à cette invasion, le scénario de mélange ayant reçu une probabilité *a posteriori* élevée (environ 0,7). Les populations sources semblent avoir des tailles efficaces élevées (de l'ordre de 10^4). Les populations envahissantes ont subi un goulot d'étranglement fort, de l'ordre d'une centaine d'individus pendant environ 5 ans (50 générations). Au total, ces résultats sont cohérents avec les données historiques et sont compatibles avec l'hypothèse que l'invasion ait commencé à partir de souches fongiques importées accidentellement via des plants de bananiers ramenés de provenances diverses au Honduras dans une collection visant l'amélioration variétale.

SUMMARY.— The reconstruction of the routes of invasions is one of the major challenges of invasion biology. The Ascomycete *Mycophaerella fijiensis*, causal agent of the Black Leaf Streak Disease of banana, stemming from South-East Asia, has started invading the American continent in 1972 and is now present in all American banana-producing regions. In a previous population genetics study at a global scale we suspected that the American invasion could have resulted from multiple introductions and admixture. In the present study, we aimed at statistically testing this hypothesis using Bayesian scenario choice based on Approximate Bayesian Computation and at inferring the demo-genetic characteristics of this introduction. To do so, we have confronted seven scenarios of disease introduction to previously acquired data consisting of molecular genotyping at 21 microsatellite markers of six American populations, two South-East Asian populations and one Oceanian population. We were able to firmly demonstrate a contribution of Oceania to this invasion. The admixture scenario received a high posterior probability (about 0.7). Parameter inferences suggest equilibrated contributions of Oceania and South-East Asia to the American invasion. The source and current invasive populations seem to have high effective population sizes (of the order of 10^4). Invasive populations went through a drastic bottleneck of the order of one hundred individuals for about 5 years (50 generations). Altogether these results are very congruent with historical data and are compatible with the assumption that the invasion started from fungal strains accidentally imported on infected banana plants brought back into Honduras from diverse origins in a collection built for the purpose of plant breeding.

Many emerging infectious diseases of plants are associated with a geographical range expansion and can be considered as biological invasions (Anderson *et al*., 2004). In such cases, understanding the factors underlying emergence requires a good knowledge the routes of introduction, i.e., the geographic pathways followed by propagules between the source and invading populations. The genetic variability available in the early stages of invasion is determined by the number and genetic composition of source populations, the demographic contribution of each source, the number of introduced individuals, the number of intermediate introduction steps between the source and the invaded region, and the dynamics of demographic expansion after each introduction (Estoup & Guillemaud, 2010; Guillemaud *et al*., 2011). Such knowledge is a prerequisite to the identification of phenotypic and genetic changes that may have been crucial to trigger emergence, as it allows precisely comparing each invasive population to its own source population(s) (Dlugosch & Parker, 2008). By deciphering the routes of invasions, it is also often possible to infer the modes of disease propagation at different geographical scales and notably distinguish between natural and human-mediated propagation. Altogether such studies provide important information for the design and optimization of prevention, quarantine, and management strategies of emerging infectious plant diseases (Estoup & Guillemaud, 2010; Barrès *et al*., 2012).

The inference of invasion history has long relied on a fine interpretation of descriptors of population genetic structure such as diversities and differentiation indices. But in the last decade, a new set of methodologies has emerged that, in principle, allows confronting concurrent complex demographic and evolutionary scenarios to genetic data within a statistically grounded framework. These are based on Approximate Bayesian Computation (ABC, Beaumont *et al.*, 2002). In ABC the limiting step of very complex likelihood computations is replaced by an approximation of the likelihood, obtained by simulating data sets under considered scenarios and selecting the simulated data sets that are closest to the observed data on summary statistics (Beaumont *et al*., 2002). It has already been successfully used to reconstruct the introduction pathways of several invasive species (e.g., Miller *et al.*, 2005; Guillemaud *et al.*, 2011) among which some plant pathogens (Barrès *et al.*, 2012, Dutech *et al.*, 2012). Both the Bayesian nature of ABC and the approximation at the core of the technique require that it should be used with caution and properly validated (Bertorelle *et al.*, 2010; Cornuet *et al*., 2010). Validation of ABC analyses may rely on two practices. The first is to test the robustness and precision on the method using datasets produced by simulation with known scenarios and parameter values (the so-called pseudo-observed data sets or PODs, Cornuet *et al.*, 2008, 2010, 2014). The second is to replicate the analysis on several independent population samples of both invaded and source areas (Lombaert *et al*., 2015).

Here, we propose to illustrate these approaches on the continental invasion of Central America by the devastating fungal plant pathogen *Mycosphaerella fijiensis*, the causal agent of the Black Leaf Streak Disease of banana (BLSD also known as Black Sigatoka; Rhodes, 1964). BLSD is the most economically important disease of bananas (Jones, 2000) and an important food-security threat worldwide (Pennisi, 2010). While its host plant (genus *Musa*) has been established across the tropics and subtropics for several centuries (Perrier *et al.*, 2011), the emergence of *M. fijiensis* only started in the 60's, with a first mention in Fiji in 1963 (Rhodes, 1964) and a rapid propagation in the Pacific Islands. In hindsight it appeared that the disease had already observed in the early 1950's in New Guinea and Solomon Islands, which were consequently proposed as the likely centre of origin of the disease (Stover, 1978). Later, population genetics studies confirmed that the centre of origin includes this area, but they also suggested that it could be much wider (including Indonesia and South China Sea countries, Carlier et al., 1996; Rivas *et al.*, 2000; Robert *et al.*, 2012). It was first described in Africa and America in the early 70's, and from there invaded the whole inter-tropical belt in less than 40 years. Historical observations and hypotheses about the outbreaks and propagation of the disease around the world are available (reviewed in Jones, 2000),

but do not by themselves constitute reliable indications to track routes of introduction (Robert *et al.*, in revision).

On the American continent, the disease was first detected during important epidemics in Honduras in 1972, but it could have been there several years before under a "quiescent" form (Stover & Dickson, 1976; Stover, 1980). These initial epidemics have occurred in a collection of banana varieties comprising samples imported since 1960 from various countries of South-East Asia to improve commercial banana varieties (Rosales *et al.*, 1999). From this point, historical records suggest that the disease has progressed gradually from country to country (Mourichon & Fullerton, 1990). This was confirmed by population genetic analyses showing an absence of clear genetic structure within countries and a large-scale pattern of isolation by distance between countries (Halkett *et al.*, 2010; Robert *et al.*, 2012). In a recent population genetics study, relying on a worldwide collection of pathogen samples (25 populations and isolated strains from 37 countries), and both microsatellite and sequence markers, Robert *et al.* (2012) found that the American invasion was genetically related to both populations from South-East Asia (Malaysia, Indonesia, Papua New Guinea, Philippines, Taiwan), as well as to invasive Oceania populations (Fiji, New Caledonia, Tonga). In a Bayesian clustering analysis using STRUCTURE (Pritchard *et al*., 2000), it was consistently found that out of 10 independent runs, two suggested a single origin likely in Oceania and eight suggested a possible admixture between Oceania and South-East Asia. In addition to this uncertainty about the identity and contribution of possible source populations, neither the intensity of bottlenecks at introduction nor the current effective population size is presently known.

In the present study, we conduct a systematic ABC scenario choice analysis to address the following questions: (i) Does the American invasion stem from an admixture between Oceania and South-East Asia populations? (ii) Have bottlenecks accompanied this introduction and, if yes, may we infer their intensity?

MATERIALS AND METHODS

STUDIED MODEL

Mycosphaerella fijiensis is a haploid heterothallic fungus displaying both an asexual reproduction and a regular sexual reproduction at almost all disease cycles (sexual spores called ascospores are produced after 30 to 40 days in favourable conditions, Jones, 2000; Churchill, 2011). The number of sexual generations each year is thus about 10. In installed epidemics, sexual reproduction is frequent enough for populations to be at Hardy-Weinberg equilibrium at the plantation scale (Carlier *et al*., 1996; Rivas *et al*., 2000; Halkett *et al*., 2010). As spores (conidia) produced asexually can disperse over scales of a few metres only, wind-dispersed ascospores are considered a more important source of inoculum for the spatial spread of the disease as they might be carried over several hundreds of kilometres (Jones, 2000; Rieux *et al*., 2014). Anthropogenic activities may also accelerate the dispersal of the disease by the transport of infected plant material (planting or packing material).

DATA

We used part of the data set analysed by Robert *et al.* (2012) (Tab. I, Fig. 1). In brief, samples were derived from a historical laboratory collection preserved in glycerol at -80°C. Each individual corresponded to a strain derived from a single ascospore isolated from an infected banana leaf. Population samples were composed of individuals sampled simultaneously on several leaves of neighbour banana trees from the same plantation. Most banana hosts were susceptible clones belonging to the AAA and AAB genome groups. DNA extraction from growing mycelium was carried out as described in Halkett *et al.* (2010). Individuals were genotyped at 21 microsatellite markers described in Neu *et al.* (1999), Zapater *et al.* (2008) and Robert *et al.* (2010). In the present analysis we included two samples from Honduras respectively collected in 1999 and 2009 and one sample from each of the following countries Costa Rica, Colombia, Panama and Venezuela. Oceanian islands were represented by the only sample of sufficient size available in our collection: Fiji. South-East Asia was represented by Philippines and Malaysia. Ideally it would have been desirable to include a sample from Papua New Guinea, as sequence-based analyses suggested a genetic proximity with American samples (Robert *et al.*, 2012). But the only sample available is in fact composed of individuals collected in various localities and cannot be

considered as a population sample. It is thus unsuitable for the type of ABC analyses that we applied. Basic population genetic properties of the studied populations are available in Robert *et al.* (2012).

TABLE I

Characteristics of the samples included in the study. H_E stands for unbiased estimate of gene diversity (Nei 1978) and A_r for allelic richness (computed using the rarefaction method implemented in HP-RARE, Kalinowski 2005)

Geographic area	Country	Acronym	Sample size	Sampling date	Sampling locality	A_r	H_E
South-East Asia	Philippines	PHL	22	1993		4.62	0.60
	Malaysia	MYS	22	1995	Langkawi	2.14	0.32
Oceania	Fiji	FIJ	25	1996	Korovou	1.76	0.18
America	Honduras	HDN	32	1999	Yojoa. Cortes	1.90	0.34
			25	2009	La Lima. Cortes	2.30	0.44
	Colombia	COL	23	2009	La Paila	1.80	0.22
	Costa Rica	CRI	33	1999	Limon	1.90	0.29
	Venezuela	VEN	25	2009	Santa Bárbara	1.65	0.21
	Panama	PAN	32	1999	Finca Jazmin	1.86	0.26

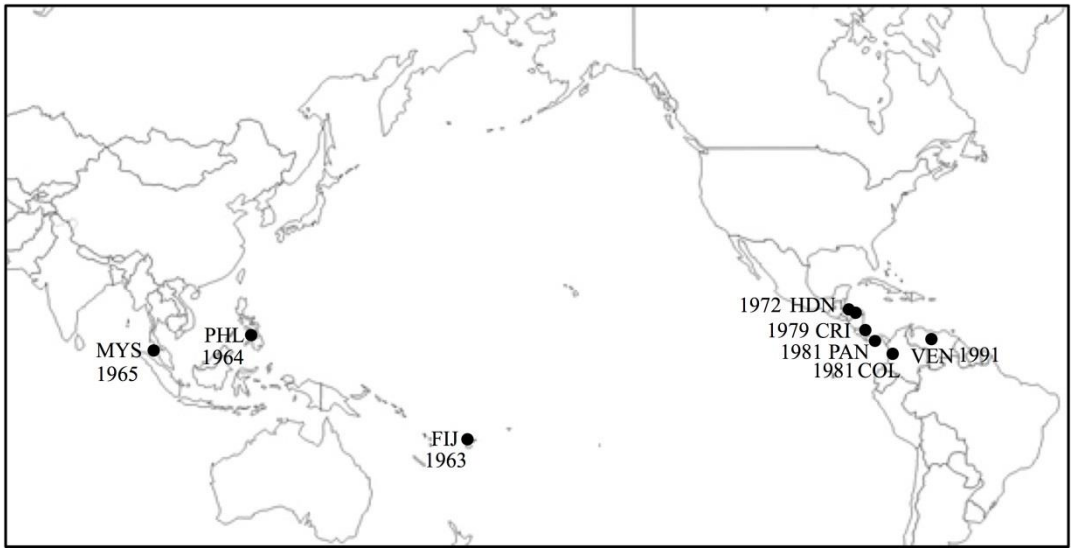

Figure 1.— Map of the samples included in the study with the corresponding dates of first report (from Jones 2000).

SCENARIOS

We designed a set of seven scenarios to test whether the American invasion stemmed from a single introduction or from multiple introductions followed by admixture and to specify the contributions of potential source populations to the invasion (Fig. 2). Scenarios 1 to 4 assume that the American sample was founded through an admixture from two source populations. In scenario 1, it is assumed that the sampled American population derives from the admixture of two sampled populations, one in South-East Asia, and the other in Oceania. As we do not have many samples from these source regions, it is possible, if not likely, that the American sample stemmed from two source populations that were not directly sampled. Scenario 2 considers the case where the South-East Asia source was not sampled. Scenario 3 assumes that the Oceania population was not sampled. Scenario 4 assumes that none of the two true source populations was sampled. Scenarios 5 to 7 assume that a single source population contributed to the invasion. In scenario 5, the American invasion stemmed from the sampled Oceania population. In scenario 6, it stemmed from the sampled South-East Asia population. In scenario 7, it was founded by an unsampled population. In all scenarios, we modelled a bottleneck at the foundation of both the

Oceanian population and the American population. The introduced populations were assumed to be isolated from each other and from the source population after the introduction with no exchange of migrants.

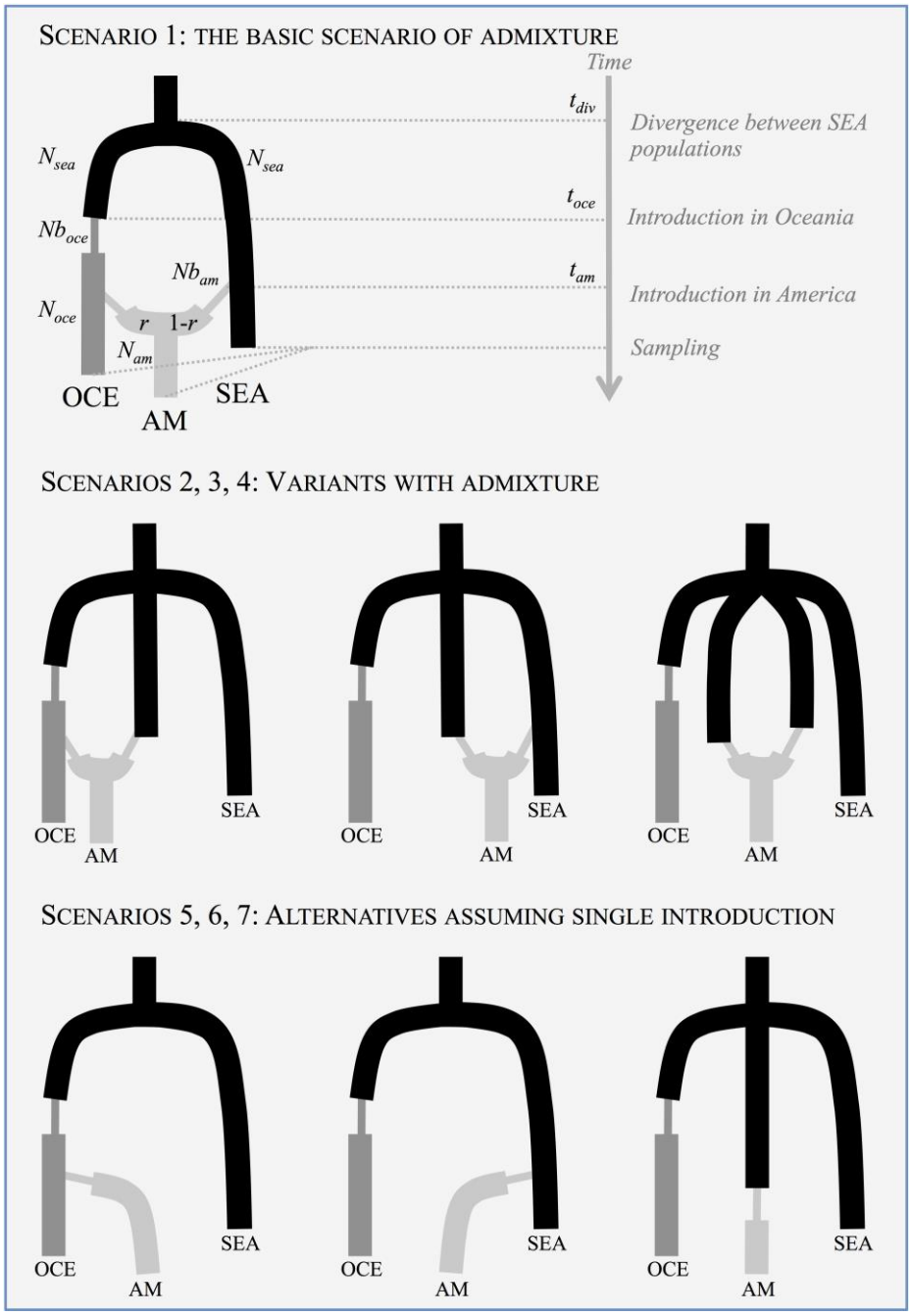

Figure 2.— The seven scenarios under comparison.

TABLE II

Prior distributions of demo-genetic parameters used in scenario choice

Parameters		Distribution	Min	Max
Contemporary effective population size				
of the American population	N_{am}	Loguniform	10^2	10^6
of the Oceanian population	N_{oce}	Loguniform	10^2	10^6
of SEA populations	N_{sea}	Loguniform	10^2	10^6
Characteristics of bottlenecks				
Effective size of the American population at foundation	Nb_{am}	Uniform	1	500
Effective size of the Oceanian population at foundation	Nb_{oce}	Uniform	1	500
Durations of bottleneck in America	db_{am}	Uniform	0	200
Duration of bottleneck in Oceania	db_{oce}	Uniform	0	200
Dates of divergence				
Date of foundation of the American population	t_{am}	Uniform	270*	340*
Date of foundation of the Oceanian population	t_{oce}	Uniform	350*	1000
Date of the divergence of South-East Asian populations	t_{div}	Loguniform	10^3	10^6
Rate of admixture	r	Uniform	0.001	0.999
Mutational parameters of microsatellite markers				
Mean mutation rate	μ	Uniform	8×10^{-6}	5×10^{-5}
Mean parameter of the geometric distribution of the number of repeats	P	Uniform	0.1	0.4

*Dates of population foundation are those used for analyses including American samples from 1999 (early sample from Honduras, Costa Rica and Panama). For analyses including samples from 2009 (late sample from Honduras, Colombia and Venezuela), 100 generations were added to these values.

TABLE III

Posterior probabilities of the seven scenarios tested. Confidence intervals are given into brackets. In bold are highlighted the highest posterior probabilities for each of the nine replicate tests

Samples included		Scenario						
		with multiple introductions and admixture				with single introduction		
		1	**2**	3	4	5	6	7
SEA = PHL	HND 1999	0.0328 [0.0302,0.0354]	**0.8634 [0.8551,0.8718]**	0.0018 [0.0012,0.0025]	0.0112 [0.0097,0.0126]	0.0791 [0.0721,0.0861]	0.0001 [0.0000,0.0007]	0.0115 [0.0100,0.0131]
	HND 2009	0.1245 [0.1106,0.1385]	**0.7432 [0.7194,0.7671]**	0.0010 [0.0008,0.0013]	0.0069 [0.0055,0.0083]	0.1183 [0.0989,0.1378]	0.0000 [0.0000,0.0001]	0.0059 [0.0047,0.0072]
	CRI	0.0517 [0.0483,0.0551]	**0.7976 [0.7883,0.8069]**	0.0028 [0.0016,0.0040]	0.0139 [0.0120,0.0158]	0.1200 [0.1119,0.1280]	0.0001 [0.0000,0.0013]	0.0139 [0.0119,0.0158]
	PAN	0.0585 [0.0282,0.0889]	**0.6867 [0.6719,0.7015]**	0.0031 [0.0000,0.0360]	0.0128 [0.0000,0.0452]	0.2251 [0.1889,0.2612]	0.0001 [0.0000,0.0332]	0.0137 [0.0000,0.0461]
	COL	0.1693 [0.1539,0.1846]	**0.5740 [0.5465,0.6015]**	0.0005 [0.0004,0.0006]	0.0031 [0.0024,0.0037]	0.2507 [0.2230,0.2783]	0.0000 [0.0000,0.0000]	0.0025 [0.0019,0.0030]
	VEN	0.2515 [0.2291,0.2739]	**0.6058 [0.5790,0.6327]**	0.0048 [0.0038,0.0058]	0.0168 [0.0138,0.0199]	0.1049 [0.0884,0.1213]	0.0005 [0.0003,0.0006]	0.0157 [0.0125,0.0189]
SEA = MYS	HND 1999	0.0524 [0.0494,0.0554]	**0.8341 [0.8280,0.8402]**	0.0004 [0.0000,0.0010]	0.0064 [0.0054,0.0073]	0.1022 [0.0970,0.1073]	0.0000 [0.0000,0.0007]	0.0045 [0.0037,0.0053]
	HND 2009	0.0224 [0.0175,0.0272]	**0.7983 [0.7700,0.8265]**	0.0020 [0.0013,0.0027]	0.0489 [0.0374,0.0605]	0.0784 [0.0599,0.0969]	0.0000 [0.0000,0.0000]	0.0500 [0.0378,0.0622]
	CRI	0.0970 [0.0924,0.1015]	**0.7370 [0.7294,0.7446]**	0.0006 [0.0000,0.0020]	0.0065 [0.0050,0.0080]	0.1546 [0.1483,0.1609]	0.0000 [0.0000,0.0014]	0.0043 [0.0029,0.0057]
	PAN	0.0835 [0.0688.0.0981]	**0.6456 [0.6368,0.6543]**	0.0007 [0.0000,0.0171]	0.0075 [0.0000,0.0237]	0.2575 [0.2393,0.2757]	0.0000 [0.0000,0.0164]	0.0053 [0.0000,0.0215]
	COL	0.0625 [0.0513.0.0737]	**0.6771 [0.6402,0.7140]**	0.0013 [0.0009,0.0017]	0.0228 [0.0177,0.0279]	0.2093 [0.1742,0.2443]	0.0000 [0.0000,0.0000]	0.0270 [0.0206,0.0335]
	VEN	0.0703 [0.0574.0.0833]	**0.6677 [0.6321,0.7034]**	0.0087 [0.0063,0.0112]	0.0891 [0.0716,0.1065]	0.0654 [0.0516,0.0792]	0.0000 [0.0000,0.0000]	0.0988 [0.0782,0.1194]

ABC IMPLEMENTATION

Twelve independent ABC scenario choice analyses were conducted using different combinations of the populations available in each of the three origins (America, Oceania, South-East Asia, Tab. III) as recommended by Lombaert *et al.* (2015). For each of these 12 replicate analyses, all usual steps of ABC were conducted using the software DIYABC 2.1.0

(Cornuet *et al.*, 2008, 2010, 2014). For each sample triplet, one million simulated data sets were produced using coalescence algorithms implemented in DIYABC. Simulations started with the latest sampling, i.e., 1999 for the six analyses including the early sample from Honduras, Costa Rica, and Panama, and 2009 for the six analyses with the late sample from Honduras, Colombia, and Venezuela.

PRIOR DISTRIBUTIONS OF PARAMETERS

Parameters were drawn in the prior distributions described in Table II. Priors on effective sizes were chosen large and uninformative because knowledge about the population biology of *M. fijiensis* is scarce. We used a generalized stepwise mutation model to simulate mutations at microsatellite markers. Each locus *l* was affected a specific mutation rate μ_l and a specific coefficient of the geometric distribution of mutation length P_l. These locus-specific parameters were drawn from gamma distributions with shape 2 and means μ and P, themselves drawn from uniform priors. Dates for the American and Oceanian population foundations were drawn from uniform priors. We considered that the American population was founded between 1965 and 1972. We also considered that Fiji was founded before 1964, but not before the early XXth century.

For each of the seven scenarios, prior choices were validated using a principal component analysis (PCA) in the space of summary statistics considering 1000 random simulated datasets (i.e., with parameter values drawn from the priors). Then the observed data set was projected as supplementary individual and it was visually checked that if fell well within the variability of simulated data (Fig. 3).

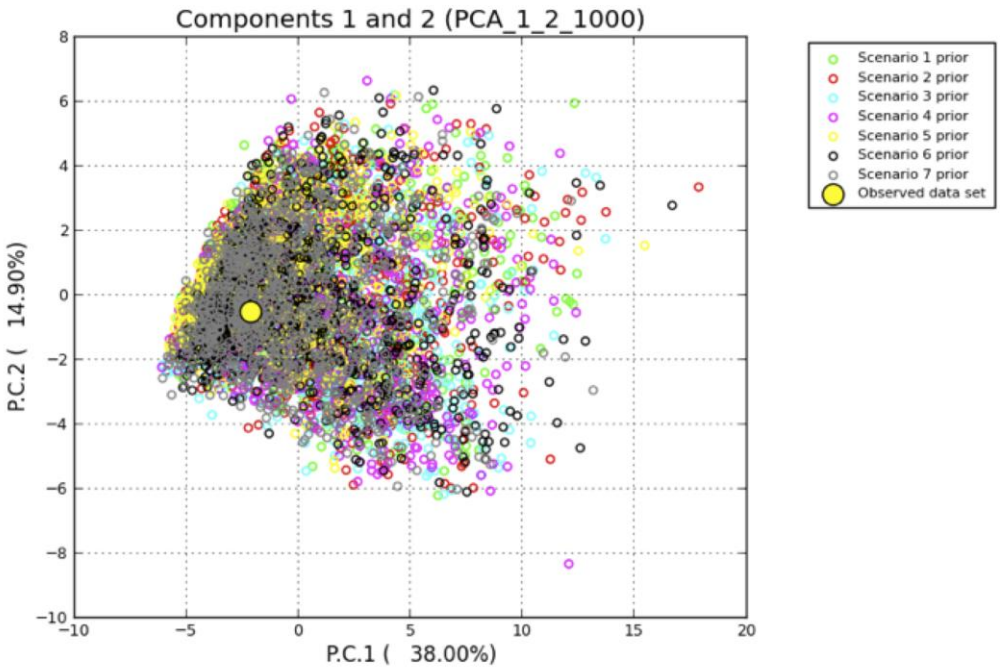

Figure 3.— Validation of prior choices.

SUMMARY STATISTICS

The similarity between real and simulated datasets was measured using 24 summary statistics: the mean number of alleles, mean expected heterozygosity, and mean allelic size variance of each sampled population and of each pair of populations, pairwise FST and shared allele distance between all pairs of populations. Other statistics available in DIYABC were left aside for later use in model checking.

POSTERIOR PROBABILITIES AND DISTRIBUTIONS

Posterior probabilities of each scenario were computed by performing a polychotomous weighted logistic regression on the 1% of simulated datasets closest to the observed dataset (Cornuet *et al.*, 2008, 2010) after linear discriminant analysis on summary statistics (Cornuet *et al.*, 2014).

After scenario choice, we proceeded to parameter inference under the most probable scenario. Posterior distributions of parameters were evaluated under this scenario using a local linear regression on the 1 % closest simulated data sets with a logit transformation. The medians of posterior distributions were used as point estimates.

ABC VALIDATION

Confidence in scenario choice was tested using additional simulations. Specifically, 500 pseudo-observed data sets (PODs) were simulated under each of the rejected scenarios drawing parameters into prior distributions. Their posterior probabilities were computed as described above. The type II error rate was then computed as the proportion of PODs for which the posterior probability is highest for the retained scenario, although it is not the one that has served in generating the data (Cornuet *et al.*, 2008).

The precision of parameter inference was then assessed using 500 pseudo-observed data sets with known parameter values drawn from the prior distributions. Model checking, i.e., the assessment of the goodness-of-fit of the obtained scenario and parameter values was performed using 500 PODs simulated using the posterior distributions of parameters. For each summary statistics available in DIYABC and not already used in parameter inference (Mean Garza-Williamson's M index, Maximum Likelihood coefficient of admixture, $(d\mu)^2$ distance between samples), the observed value was ranked against the distribution of the corresponding statistics in the PODs. We found that all the statistics tested fell well within the posterior distribution when applying the model checking to the obtained scenario (data not shown).

RESULTS AND DISCUSSION

Posterior probabilities of the seven scenarios obtained with each of the 12 replicate sets of populations are given in Table III.

A CLEAR OCEANIAN CONTRIBUTION TO THE AMERICAN INVASION

A first observation is that four scenarios (3, 4, 6 and 7) received very low statistical support. These are all scenarios with no explicit contribution of the sampled Oceanian population: Fiji. We can thus exclude the possibility that the American invasion was only caused by a South-East Asia population. In other words, there has been a non-negligible contribution of Oceania in this invasion. Although scenarios accounting for the possibility that this contribution was made by other populations than Fiji (scenario 4) were also rejected, it would be premature to conclude than Fiji directly contributed to the invasion. A proper test of this hypothesis would require a specifically designed ABC test with other Oceanian samples to compare Fiji with.

THE PROBABILITY OF ADMIXTURE

In all 12 replicate analyses, scenario 2 received the highest posterior probability ranging from 0.57 (95 % CI: 0.54-0.60) to 0.86 (95 % CI: 0.85-0.87). Scenario 2 describes the American invasion as the result of an admixture between Fiji and an unsampled South-East Asian population. Two other scenarios were associated with non-negligible posterior probabilities. Scenario 1, which assumes that American invasion started following an admixture between Fiji and the sampled South-East Asia population, obtained posterior probabilities up to 0.25 (95 % CI: 0.23-0.27) with the triplet Venezuela, Fiji and Philippines. Lastly, scenario 5, which assumes that the American invasion was founded solely by Oceania, received posterior probabilities ranging from 0.07 (95 % CI: 0.05-0.07) to 0.25 (95 % CI: 0.24-0.28). Table IV details the errors made by this scenario choice analysis for the first replicate (early sample from Honduras, Fiji and Philippines). The overall type II error rate associated with scenario 2 was low (0.047). It is

maximal for scenarios 5 (0.104) and 1 (0.076). Over the 12 replicate analyses, the inferred rate of admixture laid between 0.4 (95 % CI: 0.24-0.87) and 0.62 (95 % CI: 0.27-0.89).

TABLE IV

Confidence in scenario choice for the first replicate test (samples HND99, FIJ and PHL). For each scenario, 500 pseudo-observed data sets were produced by simulations, and their scenario was then inferred using the ABC procedure described in the text. The type II error associated with scenario 2 is the proportion of PODs simulated under scenarios other than 2, for which scenario 2 received the highest posterior probability. The global type II error of scenario 2 is 0.048

		Scenario with highest posterior probability							Type II error of scenario 2
		1	2	3	4	5	6	7	
Scenario used to generate PODs	1	327	38	25	1	58	43	8	0.076
	2	46	267	2	45	86	3	51	-
	3	33	7	228	71	0	111	50	0.014
	4	7	20	49	285	3	25	111	0.04
	5	16	52	0	4	426	1	1	0.104
	6	16	3	43	9	1	407	21	0.006
	7	11	23	30	113	2	17	304	0.046

Altogether, our results strongly support the hypothesis that the American invasion was initiated following multiple introductions and admixture. Although our sampling of potential source areas is far from complete, the analysis suggests that the source populations are related to Fiji on the one hand and Philippines on the other hand. Malaysia was excluded as a possible source population.

STRONG BOTTLENECKS AT INTRODUCTION

Estimates of effective population size under scenario 2 were associated with low precision (Tab. V), but were rather constant across the 12 replicate tests (only the first is shown). Contemporary effective population sizes were of the order of 50 000 for South-East Asian populations and 10 000 for the Oceanian invasive populations. These values are remarkably concordant with estimates obtained in the African continent using both DIYABC and MSVAR. They also are congruent with effective population size estimates obtained for the fungal pathogen of Hevea *Microcyclus ulei* in Latin America (Barrès *et al.*, 2012) and much higher than those of most other eukaryotes (Charlesworth, 2009). In all tests, the introduction in America was associated with a bottleneck severity (i.e., the ratio of duration over effective size) of 0.4. This represents the equivalent of 65 individuals over 150 generations (around 15 years). The contemporary effective population size of American populations was associated with a very poor precision but seemed much lower than that of the Oceanian population. Possible explanations may be that these populations had not yet recovered from the original bottleneck at the time of sampling, or alternatively that they go through recurrent periods of low population size because of treatments on banana producing areas.

THE HISTORY OF AN INVASION

Obviously, without a precise inference of parameters, these results do not allow any conclusion about the precise way the introduction occurred. They are intriguingly congruent with the assumption that the invasion was initiated in a Honduran banana collection.

TABLE V

Parameter inference, bias and precision of the median of posterior distributions as parameter estimate. The samples used were HND99, FIJ and PHL and the scenario considered was scenario 2. As compared to scenario choice analysis, two million simulations were conducted and priors were slightly modified with (i) bottleneck effective size ranging from 1 to 200 and bottleneck durations ranging from 0 to 200 generations (ii) t_{am} drawn in a lognormal distribution with a mode at 270 and (iii) t_{oce} drawn in a lognormal distribution with a mode at 360

	Inference			Confidence interval		Bias and precision				
Parameter	Mean	Median	Mode	q025	q975	Mean relative bias	Median relative bias	Square root of mean square error/true value	Median of the absolute error/true value	Factor 2
N_{am} μ	0.095	0.013	1.47E-3	1.39E-3	0.775	1.588	-0.081	5.603	0.899	0.282
N_{oce} μ	0.651	0.198	0.0267	0.014	3.820	0.556	-0.214	2.288	0.833	0.332
N_{sea} μ	2.000	1.940	1.650	0.892	3.420	0.020	-0.044	0.354	0.214	0.962
t_{am} μ	3.79E-3	2.72E-3	1.68E-3	1.07E-3	0.013	0.225	0.023	1.032	0.405	0.728
db_{am} μ	9.06E-4	3.87E-4	1.13E-5	1.03E-5	5.01E-3	IR	-0.275	IR	0.781	0.396
Nb_{am} μ	3.70E-3	3.35E-3	2.16E-3	5.45E-4	8.36E-3	0.423	-0.121	1.961	0.481	0.612
t_{oce} μ	0.018	0.016	0.013	4.30E-3	0.041	0.216	-0.017	0.846	0.390	0.744
db_{oce} μ	5.09E-3	5.09E-3	6.44E-3	4.23E-4	9.81E-3	0.758	-0.319	5.067	0.797	0.362
Nb_{oce} μ	3.61E-3	3.26E-3	1.96E-3	4.81E-4	8.33E-3	0.727	-0.049	2.937	0.522	0.576
t_{div} μ	0.24	0.199	0.147	0.0545	0.652	0.097	-0.080	0.720	0.299	0.826
μ	3.21E-5	3.22E-5	3.25E-5	1.35E-5	4.9E-5	0.1152	0.025	0.505	0.266	0.912
P	0.272	0.264	0.257	0.120	0.462	0.0280	-0.050	0.346	0.221	0.974
N_{am}	7460	1240	163	132	60600	1.1755	-0.106	3.323	0.901	0.276
N_{oce}	20000	7190	1260	439	92300	0.4545	-0.162	1.922	0.803	0.350
N_{sea}	60900	58800	55600	31900	96200	0.0707	0.020	0.513	0.269	0.860
t_{am}	178	147	103	102	436	0.0661	0.006	0.484	0.319	0.862
db_{am}	65.9	63.5	63.1	5.89	147	IR	-0.019	IR	0.359	0.762
Nb_{am}	134	142	176	32.9	197	0.4554	-0.029	1.758	0.335	0.762
r	0.62	0.623	0.606	0.267	0.886	0.8064	0.006	7.412	0.189	0.854
t_{oce}	544	520	321	188	968	0.1744	0.024	0.622	0.316	0.880
db_{oce}	109	111	128	10.8	195	0.7932	-0.014	3.033	0.356	0.746
Nb_{oce}	96.8	93	85	15.4	191	0.7168	-0.011	2.890	0.327	0.752
t_{div}	9470	7110	3570	1470	31200	0.0922	-0.015	0.619	0.343	0.794

Banana varieties used in such plant-breeding program are usually sought for in the centre of origin and domestication of *Musa*, which is South-East Asia. Stover (1978) also described frequent movements of plant material from Pacific to the American continent via Hawaii at that time. Historical data suggest that the disease was present in the Honduran banana collection a few years before the start of invasion in 1972 (Stover & Dickson, 1976). This lag time, which is frequently observed in invasions, nicely fits the duration of bottlenecks inferred here. We currently have no information regarding the causes of this lag. At low density, disease persistence mainly relies on the production of asexual spores, called conidia, which are known to disperse at very low distance (a few meters away, Rieux *et al.*, 2014). The species being haploid and heterothallic, the

production of sexual spores, able to disperse over much longer distances (several hundred meters on average, Rieux *et al.*, 2014), requires that two necrotic lesions of compatible sexual signs (one + and one –) lie sufficiently close to each other on the same banana leaf to exchange gametes. Sexual reproduction, and thus efficient dispersal, therefore requires high densities. The lag time may thus well correspond to the time necessary for this threshold density to be met. A non-exclusive hypothesis is that some genetic changes may have been necessary for the disease to adapt to the environmental context where it was introduced.

FUTURE PROSPECTS

The present study suggests that the American invasion stemmed from an admixture between Oceanian and South-East Asian populations. It adds to other population genetic inference studies showing that the African invasion was founded following a single introduction from one of the populations of South China Sea (Malaysia, Philippines, Taiwan, Robert *et al.*, in revision). Altogether these results pave the way to more efficient studies of potential adaptive phenotypic and genetic changes underlying the global success of *M. fijiensis* by allowing comparing American and African invasive populations to their corresponding source populations.

ACKNOWLEDGEMENTS

This paper would have never been written without the essential intellectual and practical support of Benoît Facon. Data were produced through the molecular genetic analysis technical facilities of the IFR119 with the help of E. Desmarais and F. Cerqueira. This work was supported by the Région Languedoc-Roussillon, the Agence Nationale de la Recherche (project EMERFUNDIS ANR 07-BDIV-003 and EMILE ANR 09-BLAN-0145-01) and French Agropolis Fondation (Labex Agro – Montpellier. BIOFIS project number 1001-001).

REFERENCES

ANDERSON, P.K., CUNNINGHAM, A.A., PATEL, N.G., MORALES, F.J., EPSTEIN, P.R., & DASZAK, P. (2004).— Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *TREE*, 19: 535-544.

BARRES, B., CARLIER, J., SEGUIN, M., FENOUILLET, C., CILAS, C., & RAVIGNE, V. (2012).— Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity*, 109: 1-11.

BEAUMONT, M.A., ZHANG, W., & BALDING, D.J. (2002).— Approximate Bayesian computation in population genetics. *Genetics*, 162: 2025-2035.

BERTORELLE, G., BENAZZO, A., & MONA, S. (2010).— ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.*, 19: 2609-2625.

CARLIER, J., LEBRUN, M.-H., ZAPATER, M.-F., DUBOIS, C., & MOURICHON, X. (1996).— Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. *Mol. Ecol.*, 5: 499-510.

CHARLESWORTH, B. (2009).— Effective population size and patterns of molecular evolution and variation. *Nature Rev. Genetics*, 10: 195-205.

CHURCHILL, A.C.L. (2011).— *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Mol. Plant Pathol.*, 12: 307-328.

CORNUET, J.-M., PUDLO, P., VEYSSIER, J., DEHNE-GARCIA, A., GAUTIER, M., LEBLOIS, R., MARIN, J.-M. & A.E. (2014).— DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30: 1187-1189.

CORNUET, J.-M., RAVIGNE, V., & ESTOUP, A. (2010).— Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, 11: 401.

CORNUET, J.-M., SANTOS, F., BEAUMONT, M.A., ROBERT, C.P., MARIN, J.-M., BALDING, D.J., GUILLEMAUD, T. & ESTOUP, A. (2008).— Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics (Oxford, England)*, 24: 2713-2719.

DLUGOSCH, K.M. & PARKER, I.M. (2008).— Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.*, 17: 431-449.

DUTECH, C., BARRÈS, B., BRIDIER, J., ROBIN, C., MILGROOM, M. G., & RAVIGNÉ, V. (2012).— The Chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol. Ecol.*, 21: 3931-3946.

ESTOUP, A. & GUILLEMAUD, T. (2010).— Reconstructing routes of invasion using genetic data: why, how and so what? *Mol. Ecol.*, 19: 4113-4130.

GUILLEMAUD, T., CIOSI, M., LOMBAERT, E. & ESTOUP, A. (2011).— Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics. *CR Biologies*, 334: 237-246.

HALKETT, F., COSTE, D., PLATERO, G.G.R. *ET AL.* (2010).— Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Mol. Ecol.*, 19: 3909-3923.

JONES, D.R. (2000).— Introduction to banana, abaca and enset. *In*: *Diseases of banana, abaca and enset*. CABI Publishing, Wallingford, UK.

KALINOWSKI, S.T. (2005).— HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes*, 5: 187-189.

LOMBAERT, E., GUILLEMAUD, T., LUNDGREN, J., KOCH, R., FACON, B., GREZ, A., LOOMANS, A., MALAUSA, M., NEDVED, O., RHULE, E., STAVERLOKK, A., STEENBERG, T. & ESTOUP, A. (2015).— Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Mol. Ecol.*, 23: 5979-5997.

MILLER, N., ESTOUP, A., TOEPFER, S., BOURGUET, D., LAPCHIN, L., DERRIDJ, S., KIM, K. S., REYNAUD, P., FURLAN, L., & GUILLEMAUD, T. (2005).— Multiple transatlantic introductions of the western corn rootworm. *Science*, 310: 992.

MOURICHON, X. & FULLERTON, R.A. (1990).— Geographical distribution of the two species *Mycosphaerella musicola* Leach (*Cercospora musae*) and *M. fijiensis* Morelet (*C. fijiensis*), respectively agents of Sigatoka disease and Black Leaf Streak disease in bananas and plantains. *Fruits*, 45: 213-218.

NEI, M. (1978).— Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89: 583-590.

NEU, C., KAEMMER, D., KAHL, G., FISCHER, D. & WEISING, K. (1999).— Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Mol. Ecol.*, 8: 523-525.

PENNISI, E. (2010).— Armed and dangerous. *Science*, 327: 804.

PERRIER, X., DE LANGHE, E., DONOHUE, M. ET AL. (2011).— Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *PNAS USA.*, 108: 11311-11318.

PRITCHARD, J.K., STEPHENS, M. & DONNELLY, P. (2000).— Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.

RHODES, P.L. (1964).— A new banana disease in Fiji. *Commonwealth Phytopath. News*, 10: 38-41.

RIEUX, A., SOUBEYRAND, S., BONNOT, F., KLEIN, E., NGANDO, J.E., MEHL, A., RAVIGNE, V., CARLIER, J. & DE LAPEYRE DE BELLAIRE, L. (2014).— Long-distance wind-dispersal of spores in *Mycosphaerella fijiensis*: Estimation of anisotropic dispersal kernels from an extensive field experiment. *PLoS ONE*, 9: e103225.

RIVAS, G.G., ZAPATER, M.-F., ABADIE, C. & CARLIER, J. (2004).— Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. *Mol. Ecol.*, 13: 471-482.

ROBERT, S., RAVIGNÉ, V., ZAPATER, M.-F., ABADIE, C. & CARLIER, J. (2012).— Contrasting introduction patterns among continents in the worldwide invasion of the banana fungal pathogen *Mycosphaerella fijiensis*. *Mol. Ecol.*, 21: 1098-1114.

ROBERT, S., VITALIS, R., RIEUX, A., ZAPATER, M.-F., CARLIER, J. & RAVIGNE, V. (In revision).— Single introduction and multiple severe bottleneck events characterize the emergence of a banana fungal disease across the African continent.

ROSALES, F., TRIPON, S. & CERNA, J. (1999).— *A tribute to the work of Paul H. Allen: A catalogue of wild and cultivated bananas.* INIBAP/CIID/EARTH, Montpellier, France.

STOVER, R.H. (1978).— Distribution and probable origin of *Mycosphaerella fijiensis* in South-East Asia. *Trop. Agric., Trinidad*, 55: 65-68.

STOVER, R.H. (1980).— Sigatoka leaf spots of bananas and plantains. *Plant Disease*, 64: 750-756.

STOVER, R.H. & DICKSON, J.D. (1976).— Banana leaf spot caused by *Mycosphaerella musicola* and *M. fijiensis* var. *difformis*: a comparison of the first Central American epidemics. *FAO Plant Prot. Bull.*, 24: 36-42.